DOCUMENT RESUME

ED 201 661 TM 810 254

AUTHOR Davison, Mark L.
TITLE Loglinear Analysis of Learning Hierarchy and Developmental Sequence Data.
SPONS AGENCY National Academy of Education, Washington, D.C.: National Inst. of Education (ED), Washington, D.C.
REPORT NO NIE-G-79-0021
PUB DATE 14 Apr 81
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Cognitive Development; *Hypothesis Testing; *Mathematical Models; *Probability; Tables (Data)
IDENTIFIERS Contingency Tables; *Hierarchical Learning; *Log Linear Models

ABSTRACT

The interest in developmental sequences and learning hierarchies is growing. One approach to the study of such sequences is to gather data on several variables, each of which corresponds to a stage, step, or phase in the sequence and to examine the associations between the variables as displayed in a contingency table. If the variables are associated in ways predicted by the hypothesized sequence, then the data lend support to the sequence. Goodman's loglinear model for developmental or learning sequences is presented and illustrated on number concept data gathered by Brainerd and Fraser. Where its strong assumptions are satisfied, the model provides a probabilistic framework within which to: (1) test the plausibility of an hypothesized developmental sequence or learning hierarchy; (2) compare several hypothesized sequences on the same data; (3) estimate the proportion of subjects who do not conform to the sequence; and (4) estimate the proportion of subjects at each step in the sequence. (Author/RL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED2016661

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Loglinear Analysis of Learning Hierarchy and Developmental Sequence Data

Mark L. Davison

University of Minnesota

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

M. L. Davison

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This research was supported by a National Academy of Education Spencer Fellowship and a grant from the National Institute of Education (NIE-G-79-0021)

Paper presented to the American Educational Research Association, Los Angeles, California, April 14, 1981.

TM 810 259

Abstract

Goodman's loglinear model for developmental or learning sequences is presented and illustrated on number concept data gathered by Brainerd and Fraser. Where its strong assumptions are satisfied, the model provides a probabilistic framework within which to (a) test the plausibility of an hypothesized developmental sequence or learning hierarchy, (b) compare several hypothesized sequences on the same data, (c) estimate the proportion of subjects who do not conform to the sequence, and (d) estimate the proportion of subjects at each step in the sequence.

Loglinear Analysis of Learning Hierarchy and Developmental Sequence Data

The interest in developmental sequences and learning hierarchies is growing. One approach to the study of such sequences is to gather data on several variables, each of which corresponds to a stage, step, or phase in the sequence and to examine the associations between the variables as displayed in a contingency table. If the variables are associated in ways predicted by the hypothesized sequence, then the data lend support to the sequence.

Our purpose is to describe and critically evaluate a class of loglinear models for contingency table data which can be used to study a priori hypotheses about developmental sequences or learning hierarchies. Interested readers can refer to earlier works by Bishop, Fienberg, and Holland (1975), and Fienberg (1977) for more details on loglinear models.

The Loglinear Model

Although the model can be extended to any desired number of variables, let us assume for convenience that there are exactly three response variables; A, B, and C; which can take on values a, b, and c respectively. The three response variables define a 3-way contingency table. Each way of the table corresponds to one of the three variables. Within a way, each level represents one value which can be taken by the corresponding variable. The frequency in cell (a,b,c) of the table would represent the number of observations scored at level a on A, b on B, and c on C.

The hypothesized sequence (or each sequence if there is more than one) is presumed to divide the contingency table cells into two sets, a set of inadmissible cells and a set of admissible cells. An admissible cell

corresponds to a pattern of scores which might be expected for someone who conforms to the hypothesized sequence. Each inadmissible cell represents a pattern which violates the sequence. Deriving the admissible cells generated by a theory is itself an important and sometimes difficult step in the explication of a theory. Davison (1979), Davison, King, Kitchener, and Parker (1980) Froman and Hubert (1980) and Wohlwill (1973) enumerate the admissible cells for various kinds of theories.

As Goodman (1975) develops the loglinear approach, subjects are divided into $\underline{K} + 1$ classes. π_o designates the population proportion of subjects in the first class, which contains persons whose development does not conform to the hypothesized sequence. This first class is called the unscaleable class. Each of the remaining classes corresponds to one of the admissible cells. For $\underline{k} = 1, \ldots, \underline{K}$, $\pi_{\underline{k}}$ represents the proportion of subjects in the population who have advanced along the sequence to the point where they should exhibit the $\underline{k}^{th}$ admissible score pattern.

For members of the unscaleable class, the response variables are presumed to be independent. Within this class, $\pi(\underline{a},\underline{b},\underline{c}) = \pi(\underline{a})\pi(\underline{b})\pi(\underline{c})$. Consequently, the joint probability of observing an individual from the unscaleable population with scores $(\underline{a},\underline{b},\underline{c})$ is $\pi_o\pi(\underline{a})\pi(\underline{b})\pi(\underline{c})$. The members of the $\underline{k}^{th}$ scaleable subpopulation are all assumed to exhibit the $\underline{k}^{th}$ admissible pattern. Consequently, the joint probability of observing an individual from the $\underline{k}^{th}$ scaleable subpopulation who exhibits pattern $(\underline{a},\underline{b},\underline{c})$ is $\pi_{\underline{k}}$ if $(\underline{a},\underline{b},\underline{c})$ is the $\underline{k}^{th}$ admissible score pattern, and it is 0 otherwise. In the total population the probability of observing pattern $(\underline{a},\underline{b},\underline{c})$ is assumed to be the sum of joint probabilities. That is, the probability of observing pattern $(\underline{a},\underline{b},\underline{c})$ can be obtained by summing the joint probability of observing $(\underline{a},\underline{b},\underline{c})$ in each of the $(K + 1)$ subpopulations.

This leads to the fundamental equation of the model:

$$\pi(A,B,C) = \pi_0\pi(A)\pi(B)\pi(C) \text{ if } (A,B,C) \text{ is inadmissible} \quad (1)$$
$$= \pi_k + \pi_0\pi(A)\pi(B)\pi(C) \text{ if } (A,B,C) \text{ is admissible.}$$

There are several algorithms for fitting the model of Equation 1 (Goodman, 1975; Davison, 1980; Bishop, Fienberg, & Holland, 1975; Fienberg, 1977) and several computer programs for implementing the algorithms (Davison & Thoma, Note 1, 1980; Dixon & Brown, 1979; Larntz, Note 2, 1974). These programs provide estimates of expected cell frequencies under the model, Pearson and likelihood ratio goodness-of-fit statistics, and estimates of quantities from which the model parameters can be obtained. Davison (1980) shows how to estimate model parameters from the output of the Davison and Thoma (Note 1, 1980) algorithm.

Given multinomial assumptions, the Pearson and likelihood ratio statistics will be distributed as chi square variables under the null hypothesis represented by Equation 1. These statistics will have $N - N_A - N_B - N_C - K + NW - 1$ degrees of freedom. Here N is the total number of cells; N_A, N_B, and N_C are the number of levels along each way of the contingency table, NW is the number of ways in the table, and K is again defined as the number of admissible cells. This brings us to a limitation of the loglinear approach as developed in Goodman (1975). If we are not to run out of degrees of freedom, then K must be smaller than $(N - N_A - N_B - N_C + NW - 1)$. The example presented below will illustrate a situation in which some of the hypothesized sequences cannot be fitted to the data, because the loglinear model for those sequences requires more degrees of freedom than the data can sustain.

Davison (1980) presents a more restricted variation of the Goodman (1975) model, a variation which can sometimes be applied when Goodman's unrestricted

formulation requires too many degrees of freedom. Davison (1980) imposes the constraint that the ratio $[1 + \pi_k / \pi_0\pi(a)\pi(b)\pi(c)]$ must equal a constant for every admissible pattern. According to this constraint, the ratio of the probability of observing an admissible pattern in the scaleable subpopulation to the probability of observing that same pattern in the unscaleable subpopulation must be roughly the same for every admissible pattern. By "roughly" the same, we mean the same except for the additive constant 1 contained in the restriction. Substantively, this means that those patterns which are most commonly found in the unscaleable subpopulation are also those most commonly found among those who conform to the hypothesized sequence. While this constraint is highly restrictive, the example below illustrates data which satisfy the restricted form of Equation 1. Other examples can be found in Davison (1979, 1980) and Davison *et al.* (1980). No matter how many admissible cells are generated by the hypothesized sequence, the Pearson and likelihood ratio fit statistics will always have $(N - N_A - N_B - N_C + NW - 2)$ degrees of freedom

If $\pi_k = 0$ for all of the admissible patterns, then the response variables satisfy the independence model. In that case, there is no need to postulate a developmental sequence to account for structure among response variables because the data do not suggest that such structure exists. The data fully support the developmental sequence model of Equation 1 only if the independence model can be rejected, suggesting there is some structure to the data, and the sequence model of Equation 1 fits the data.

If $\pi_0 = 0$, then Equation 1 is a deterministic model in which every subject's response pattern is admissible. Or in other words, the deterministic form of the hypothesized developmental sequence is a limiting case of the stochastic model in Equation 1.

Statements of sequences found in the developmental and instructional literatures are typically deterministic, without any suggestion as to how measurement and sampling error should be handled. Any probabilistic sequence model, such as Equation 1, cannot be just a straightforward restatement of the deterministic sequence hypothesized in the literature, because the stochastic model must incorporate augmenting assumptions about error to translate the deterministic sequence into probabilistic form. If the data satisfy the probabilistic model, then they lend support both to the hypothesized developmental sequence and the augmenting assumptions. If, on the other hand, the probabilistic model fails to fit the data, the failure may be because the developmental sequence is incorrect, the augmenting assumptions are not satisfied, or both. The loglinear analysis itself does not disentangle the possible sources of poor fit.

Comparisons Between Sequences

Rather than determining whether a given sequence can be said to describe the data, a researcher may be interested in comparing several sequences to decide which best describes a set of data. Within the loglinear framework, there are two possible approaches to comparing sequences. The first approach incorporates the restricted model. After fitting the restricted model for each sequence, the several sequences can be compared on the basis of their Pearson or likelihood ratio fit statistics. The several fit statistics will be comparable, because they will all correspond to models having exactly the same number of degrees of freedom, and all will be based on the same data. To our knowledge, there is no way to test the statistical significance of differences in fit for the several models.

The second approach incorporates the unrestricted form of the model in Equation 1. Within this approach, the goodness-of-fit statistics for two sequences can be directly compared only if the two sequences generate exactly the same number of admissible cells. Only then will the two fit statistics have equal degrees of freedom. Two sequences with unequal numbers of admissible cells cannot be compared directly on the basis of their fit statistics unless one sequence constitutes a special case of the other.

To see how models can be compared if one is a special case of the other, consider two sequences such that the admissible cells for Sequence I are a proper subset of those for Sequence II. Let the subscript $\underline{m} = 1, \ldots, \underline{M}$ designate those cells which are admissible according to Model II but not Model I. Given the unrestricted form of Equation 1, Model I is a special case of Model II in which $\pi_{\underline{m}} = 0$ for all $\underline{m}$, the difference between the two likelihood ratio fit

statistics for Models I and II, $G_I^2 - G_{II}^2$, is itself approximately distributed as a chi square statistic with $\underline{M}$ degrees of freedom under the null hypothesis $\underline{\pi}_m = 0$ for all $\underline{m}$ and given that responses satisfy the more general model. If the null hypothesis cannot be rejected, then the more general Model II cannot be said to significantly improve the fit. Parsimony would favor Model I.

In summary, comparisons between sequences based on goodness-of-fit statistics and the unrestricted version of Equation 1 would be limited to those cases in which the two models compared have equal degrees of freedom and those cases in which one model is a special case of the other. If the restricted form of Equation 1 is applied, any two sequences can be compared regardless of how many admissible cells each generates.

Example

Our example is based on data from Brainerd and Fraser's (1975) study of number development. Brainerd and Fraser scored each subject at one of three ordination levels and one of three cardination levels. Table 1 displays the frequency with which subjects were jointly scored at each level of ordination and cardination. Figure 1 depicts four developmental sequences which might be used to explain their data; reciprocal priority with ordination preceding cardination (A), reciprocal priority with cardination preceding ordination (B), unilateral priority (C), and synchrony (D). Hatched cells are inadmissible. Numbered cells are admissible.

The unrestricted form of Equation 1 could be fitted only for sequence D. After estimating the row and column marginals, the data in Table 1 contain only four remaining degrees of freedom. Sequences A, B, and C have either five or six admissible cells. Consequently, the unrestricted model for these sequences requires at least six or seven remaining degrees of freedom. The restricted version of Equation 1 can be and was applied to all four sequences.

Subject's level of ordination and cardination do not appear to be independent. The Pearson and likelihood ratio statistics were both statistically significant ($X(4)^2 = 17.60$, $G(4)^2 = 19.99$, $p < .01$) leading us to reject the independence model.

For each sequence in Figure 1, we then fitted the restricted version of Equation 1 using the CONSCAL program of Davison and Thoma (Note 1, 1980). Equations 11 and 13 in Davison (1980) were then used to estimate the probabilities in Equation 1 from the parameters printed by CONSCAL. Only sequence C, unilateral priority, would be rejected at any conventional significance level ($X(3)^2 = 10.82$, $G(3)^2 = 11.19$, $p < .05$) irrespective of which fit measure is employed. The two

reciprocal priority models fit equally well ($\underline{X}(3)^2 = 4.26$, $\underline{G}(3)^2 = 3.37$, $\underline{p} > .05$) and better than the synchrony model. Using a .05 level of significance, the Pearson statistic ($\underline{X}(3)^2 = 7.23$) would lead to rejection of the synchrony model. The likelihood ratio statistic would not ($\underline{G}^2 = 6.19$).

For the two models which best fit the data, the reciprocal priority sequences, Table 2 displays the estimates of model parameters. For sequence A, the parameter estimates suggest that 51% of the subjects in the population are unscaleable; that is, they are not conforming to the hypothesized sequence. Thirteen percent are found at step 1 in the sequence, 11% at step 2, 3% at steps 3 and 4, and 20% at step 5. For sequence B, parameter estimates suggest that 59% fail to conform, 13% are found at step 1, 1% at step 2, 3% at step 3, 4% at step 4, and 20% at step 5.

The parameter estimates, $\underline{\pi}_o$, strongly indicate that neither sequence A nor B can be considered a "universal" sequence, because the majority of subjects fail to conform to either sequence. Although the fit statistics for the two models are identical, A might be preferred. If A rather than B is taken to be the sequence accounting for dependencies in Table 1, then a slightly higher proportion of subjects can be said to conform. Parameter estimates suggest that very few subjects occupy intermediate steps 3 and 4 in sequence A or steps 2 through 4 in sequence B.

For any model that can be said to fit the data, the difference between the likelihood ratio statistic for the independence model and the developmental sequence model is itself approximately distributed as a chi square statistic with one degree of freedom under the null hypothesis $\underline{\pi}_1 = \underline{\pi}_2 = \underline{\pi}_3 = \underline{\pi}_4 = \underline{\pi}_5 = 0$.[1] If this conditional likelihood ratio difference statistic leads to rejection of the null hypothesis at some chosen significance level, then the sequence model

can be said to fit the data significantly better than the independence model. For both models A and B, the conditional likelihood ratio statistic ($G^2(4) - G(3)^2 = G(1)^2 = 19.99 - 3.37 = 16.62$, $p < .01$) suggests that the developmental sequence model significantly outperforms the independence model.

Discussion

The loglinear approach to the study of sequences offers several advantages over alternative approaches. Unlike the Airasian and Bart (1975) and Cliff (1979) models, the loglinear model is stochastic rather than deterministic. Whereas Dayton and MacReady's (1976) method applies only to tables having exactly two levels along each way, the present approach can be applied to tables having any number of levels along each way. Furthermore, the loglinear analysis is quite rich. It provides a basis for comparing hypothesized sequences; it provides tests of fit for each separate model; it provides estimates of the proportion falling at each step along the sequence; and it provides estimates of the proportion who fail to conform to the hypothesized sequence.

On the negative side, the assumptions of the model are strong, particularly if the restricted form of Equation 1 is used. Because the analysis relies on chi square goodness-of-fit statistics, it suffers from the problems associated with such statistics. If the degrees of freedom are small, then the statistical test has low power. Some cells may need to be collapsed if their frequencies are too small.

There are two problems which will, we suspect, complicate the study of developmental sequences via the loglinear or any other method cited above. First, the admissible cells generated by two sequences can differ by as little as one cell. When the choice of sequence depends so heavily on such a small portion of the data, large sample sizes will be needed to reliably distinguish between the sequences. When comparing sequences generating highly similar admissible sets, the sequence favored may vary inconsistently from one study to the next.

Second, whether a step in an hypothesized sequence is needed to describe

responses will depend, in part, on the developmental or instructional level of persons studied. If the subjects are not advanced, then the highest steps in a sequence may not be needed to account for the data simply because no subjects have reached those steps. Similarly, in an advanced group, the lowest steps may not be needed. Consequently, researchers investigating the same hypothesized sequence in similar populations, but at different points of instruction, may arrive at quite different conclusions, even if that sequence provides a useful description of learning in both groups. If sequences contain quite transitory steps, then at any given time, few people would be found at that step. Consistent evidence for the transitory step would be difficult to obtain.

Footnote

[1]If the unrestricted form of the model is applied, then the number of degrees of freedom for the likelihood ratio difference statistic equals the number of admissible cells.

REFERENCE NOTES

1. Davison, M. L., and Thoma, S. J. CONSCAL: A FORTRAN IV program for testing structural hypotheses. Unpublished manuscript, University of Minnesota, 1979.

2. Larntz, K. ITFIT: Iterative proportional fitting program. Unpublished computer program, University of Minnesota, Department of Applied Statistics, 1974

REFERENCES

Airasian, P. W., & Bart, W. M. Validating a priori instructional hierarchies. Journal of Educational Measurement, 1975, 12, 163-173.

Bishop, Y. M. M., Fienberg, S. E. , & Holland, P. W. Discrete multivariate analysis: Theory and practice. Cambridge, Mass: MIT Press, 1975.

Brainerd, C. J. & Fraser, M. A further test of the ordinal theory of number development. Journal of Genetic Psychology, 1975, 127, 21-33.

Cliff, N. Test theory without true scores? Psychometrika, 1979, 44, 373-393.

Davison, M. L. Testing a unidimensional, qualitative unfolding model for attitudinal or developmental data. Psychometrika, 1979, 44, 179-194.

Davison, M. L. A psychological scaling model for testing order hypotheses. British Journal of Mathematical and Statistical Psychology, 1980, 123-141.

Davison, M. L., King, P. M., Kitchener, K. S., & Parker, C. A. The stage sequence concept in cognitive and social development. Developmental Psychology, 1980, 16, 121-131.

Davison, M. L. & Thoma, S. J. CONSCAL: A FORTRAN IV program for testing structural hypotheses. Applied Psychological Measurement, 1980, 4, 8.

Dayton, C. M., & Macready, G. B. A probabilistic model for validation of behavioral hirarchies. Psychometrika, 1976, 41, 189-204.

Dixon, W. J., & Brown, M. B. (Eds.) BMDP-79: Biomedical computer programs P-series. Berkeley, CA: University of California Press, 1979.

Fienberg, S. E. The analysis of cross-classified categorical data. Cambridge, Mass: MIT Press, 1977.

Froman, T. & Hubert, L. J. Application of prediction analysis to developmental priority. Psychological Bulletin, 1980, 87, 136-146.

Goodman, L. A. A new model for scaling response patterns: An application of the quasi-independence concept. Journal of the American Statistical Association, 1975, 70, 755-768.

Wohlwill, J. The study of behavioral development. New York: Academic Press, 1975.

TABLE 1

Bivariate Frequency Distribution Between
Ordination and Number Conservation

Ordination Stage	Number Conservation Stage		
	I	II	III
I	16	3	1
II	15	3	3
III	23	4	27

TABLE 2

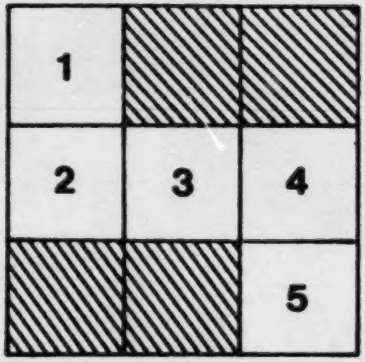
Model Parameters

Sequence	π_o	π_1	π_2	π_3	π_4	π_5	$\pi(O_I)$	$\pi(O_{II})$	$\pi(O_{III})$	$\pi(C_I)$	$\pi(C_{II})$	$\pi(C_{III})$
A	.51	.13	.11	.03	.03	.20	.15	.12	.73	.65	.16	.19
B	.59	.13	.01	.03	.04	.20	.11	.33	.56	.73	.05	.22

Figure Caption

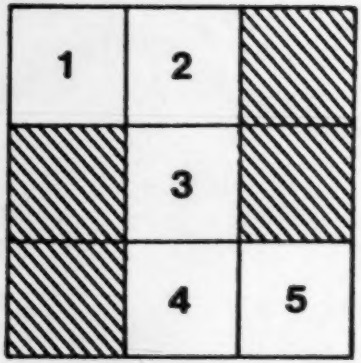
Figure 1. Admissible and inadmissible response patterns for four developmental sequences. Hatched cells are inadmissible. Numbered cells are admissible.

A



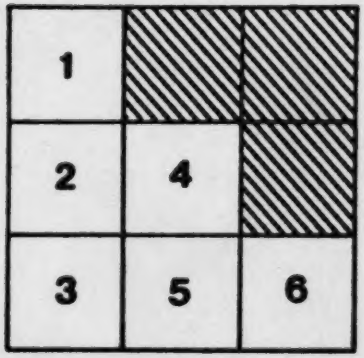
Reciprocal priority: Ordination precedes conservation

B



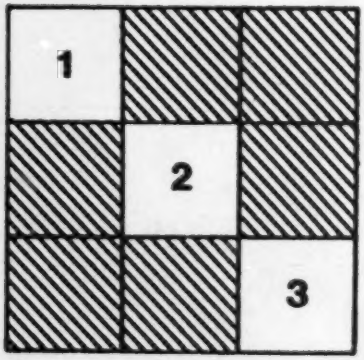
Reciprocal priority: Conservation precedes ordination

C



Unilateral priority

D



Synchrony